



Deliens, G, Antoniou, K, Clin, E, Ostashchenko, E ORCID logoORCID: <https://orcid.org/0000-0002-4817-5753> and Kissine, M (2017) Context, facial expression and prosody in irony processing. Journal of Memory and Language, 99. pp. 35-48. ISSN 0749-596X

Downloaded from: <https://e-space.mmu.ac.uk/627517/>

Version: Accepted Version

Publisher: Elsevier

DOI: <https://doi.org/10.1016/j.jml.2017.10.001>

Usage rights: Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Please cite the published version

<https://e-space.mmu.ac.uk>

Context, facial expression and prosody in irony processing

Gaétane Deliens^a, Kyriakos Antoniou^{a,c,d}, Elise Clin^{a,b}, Ekaterina Ostashchenko^a, Mikhail Kissine^{a,*}

^a*ACTE at LaDisco and Neuroscience Institute, Université libre de Bruxelles, CP 175, 50 avenue F.D. Roosevelt, 1050 Brussels, Belgium*

^b*F.R.S.-FNRS, Fund for Scientific Research, Wallonia-Brussels Federation of Belgium*

^c*Current affiliation: Department of Theoretical and Applied Linguistics, University of Cambridge, English Faculty Building, 9 West Road, Cambridge, CB3 9DP, United-Kingdom*

^d*Current affiliation: School of Humanities, Hellenic Open University, 18 Parodos Aristotelous, Patra, 26335, Greece*

Abstract

While incongruence with the background context is a powerful cue for irony, in spoken conversation ironic utterances often bear non-contextual cues, such as marked tone of voice and/or facial expression. In Experiment 1, we show that ironic prosody and facial expression can be correctly discriminated as such in a categorization task, even though the boundaries between ironic and non-ironic cues are somewhat fuzzy. However, an act-out task (Experiments 2 & 3) reveals that prosody and facial expression are considerably less reliable cues for irony comprehension than contextual incongruence. Reaction time and eye-tracking data indicate that these non-contextual cues

*Corresponding author

Email addresses: gaetane.delien@ulb.ac.be (Gaétane Deliens), ka353@cam.ac.uk (Kyriakos Antoniou), eliscclin@ulb.ac.be (Elise Clin), ekaterina.ostashchenko@ulb.ac.be (Ekaterina Ostashchenko), mkissine@ulb.ac.be (Mikhail Kissine)

entail a trade-off between accuracy and processing speed. These results suggest that interpreters privilege frugal, albeit less reliable pragmatic heuristics over costlier, but more reliable, contextual processing.

Keywords: irony; figurative language; prosody; facial expression; context; eye-tracking

1 Introduction

2 Imagine that, as you announce that you will not attend a crisis meeting
3 because of a party, your boss replies ‘I love your sense of responsibility!’.
4 Most likely, the incongruity of her comment with the conversational context
5 – broadly understood as shared background knowledge or beliefs (in the clas-
6 sic sense of [Stalnaker, 2002](#)) – will (correctly) prompt you to interpret it as
7 ironic. While such ironic utterances pervade our daily conversations, irony is
8 notoriously difficult to define in precise terms (e.g. [Gibbs, 2000](#); [Gibbs & Col-](#)
9 [ston, 2012](#), p. 52) and surfaces under many different guises (such as sarcasm,
10 jocularity, hyperbole, rhetorical question, and understatement). Neverthe-
11 less, in one sense or another, all ironically intended messages deliberately
12 mismatch the utterance literal content.¹

13 Incongruence with the background context, of the kind just illustrated,
14 is a powerful cue for irony ([Gerrig & Goldvarg, 2000](#); [Kreuz & Link, 2002](#)).
15 However, there are indications that a statement may still be interpreted as
16 ironic in the absence of such contextual incongruity, provided that other cues

¹ Of course, contextual incongruence does not necessarily boil down to manifest falsity; for instance, hyperbolic, but nevertheless literally true statements may be ironic ([Sperber & Wilson, 1981](#); [Kreuz & Roberts, 1995](#)).

17 are available (e.g. [Kowatch et al., 2013](#); [Jacob et al., 2016](#)). In particular,
18 spoken ironic utterances are often associated with a specific facial expression
19 and a distinctive prosody (e.g. [Attardo et al., 2003](#); [Rankin et al., 2009](#)). To
20 the extent that such cues to irony do not directly rely on background context,
21 in what follows we will dub them ‘non-contextual’, as opposed to contextual
22 incongruity.

23 The precise role non-contextual cues play in irony processing remains
24 ill understood. On one hand, there is some evidence that a global ironic
25 prosody can be correctly discriminated from a non-ironic one ([Bryant &
26 Fox Tree, 2005](#)), provided that the statement is uttered in a familiar language
27 ([Cheang & Pell, 2011](#)). And, in fact, many experimental designs implicitly
28 presuppose that ironic prosody is efficient, as they use a distinctive prosody to
29 contrast between ironic and literal stimuli (e.g. [Chevallier et al., 2011](#); [Colich
30 et al., 2012](#); [Kowatch et al., 2013](#)). On the other hand, [Bryant & Fox Tree
31 \(2005\)](#) report that a prosodic contour that is successfully discriminated as
32 ironic is also perceptually associated with other dimensions, such as anger
33 or inquisitiveness. Furthermore, the perception of a given prosodic contour
34 as ironic or not may be influenced by the contextual availability of an ironic
35 interpretation ([Voyer et al., 2016](#)).

36 We submit that while ironic tone of voice and/or ironic facial expression
37 may be correctly *discriminated*, these cues are not necessarily efficient in a
38 genuine process of irony *comprehension*. Arguably, successful social interac-
39 tions do not reduce to tagging statements as literal or not (viz. discrimina-
40 tion), but require the identification of the speaker’s discourse goals, and the
41 selection of an appropriate reaction (viz. comprehension; see [Kreuz 2000](#)).

42 Studies in brain-damaged patients suggest a dissociation between these two
43 processes: some patients fail to understand the speaker’s intent when contex-
44 tual and prosody cues are available, even though they are able to identify the
45 tone of voice as sarcastic (McDonald, 2000; McDonald & Pearce, 1996). Yet,
46 irony processing is usually investigated through tasks in which participants
47 have to judge as quickly as possible if statements are ironic or not, thus mea-
48 suring only the discrimination component. For instance, Bryant & Fox Tree
49 (2002) found that participants successfully discriminate ironic vs. non-ironic
50 utterances based on their prosody.² However, making decisions in a binary,
51 forced-choice task is very different from interpreting a message in the same
52 way as would its actual addressee. The precise role of prosody within irony
53 comprehension is further blurred by the fact that Bryant & Fox Tree (2002)
54 found context to be a more powerful cue for ironic judgements than prosody.

55 A notable exception to such metalinguistic decision paradigms is the
56 study by Kowatch et al. (2013), who designed an innovative ‘shopping task’
57 that positions participants as active interpreters. In this experimental design,
58 a puppet faces food items (e.g. an apple and an orange) and utters literal or
59 ironic statements about what it wants to buy (e.g. ‘I just love apples’). Only
60 the puppet’s tone of voice allows to disentangle ironic criticisms (e.g. ‘I just
61 love apples’), literal criticisms (e.g. ‘I just hate apples’) and literal praise
62 (e.g. ‘I just love oranges’). Participants are asked to put in a shopping cart
63 the food item the puppet really wants. In this way, participants’ response

² There are many other experimental studies that approach irony exclusively through discrimination; see, for instance, Kreuz & Roberts (1995); Climie & Pexman (2008); Epley et al. (2004); Chevallier et al. (2011); Colich et al. (2012).

mirrors their interpretation of the discourse goals of the speaker. The results of Kowatch et al. (2013) display an interesting asymmetry between accuracy and reaction time. The rate of correct responses for ironic items is low (less than 60%), and significantly so relative to literal items. At the same time, the authors report no difference in processing time or in frequencies of first looks to the correct object for ironic and literal criticisms. It could be the case, then, that while ironic prosody and/or facial expression are not very reliable for accurately grasping an ironic communicative intention, they still prompt a rapid, cognitively shallow attribution of ironic intentions to the speaker.

Importantly, Kowatch et al. (2013) did not compare ironic prosody relative to the role of context, so it is unclear whether interpreters still use prosody when context is available, and if yes, whether non-contextual cues merely complement context-based processing or whether they may take precedence over it. There is ample evidence that mastery of irony presupposes complex mental-state attribution skills (e.g. Akimoto et al., 2012; Bryant, 2012; Spotorno & Noveck, 2014). Such mentalising processes require inferring the speaker's intention by assessing the utterance content against the background context. Some theorists hold that any type of pragmatic processing involves complex, context-based inferences about the speaker's communicative intentions (Sperber & Wilson, 2002). Consistently with this idea, in 5- to 7-year-old children, it is the capacity to attribute multilayered mental states, and not ironic prosody, that predicts correct discrimination between irony and white lies (Wimmer & Leekam, 1991; see also Filippova & Astington, 2010).

89 However, it is also plausible that conversationally experienced interpreters
90 sometimes rely more on salient non-contextual cues than on context. For in-
91 stance, [Deliens et al. \(2017\)](#) recently found that in the presence of salient
92 ironic prosody, participants do not engage in context-based perspective-
93 shifting to gauge the sarcastic nature of a message. According to the parallel-
94 constraint-satisfaction account ([Katz, 2005](#); [Pexman, 2008](#)), all cues are pro-
95 cessed in parallel and activate a certain – possibly ironic – interpretation.
96 However, as acknowledged by [Pexman \(2008\)](#) herself, this model does not
97 currently provide any indication as to the relative weight of different cues.
98 A more radical idea, to which we subscribe, is that the presence of salient,
99 albeit perhaps less reliable, non-contextual cues prompts interpreters to dis-
100 regard costlier contextual processing. This hypothesis is consistent with the
101 Direct Access view (e.g. [Gibbs, 2002](#)), which predicts that interpreters do
102 not always need to analyse literal meaning in full to form a hypothesis about
103 the meaning communicated by the speaker. It is also in line with a model
104 of pragmatics according to which interpreters are driven by considerations
105 of cognitive economy, and do not necessarily engage in extensive context-
106 driven reasoning about speaker’s intentions ([Kissine, 2016](#); see also [Ferreira](#)
107 [& Patson, 2007](#)).

108 By contrast, Giora’s Graded Salience theory ([Giora, 2003](#); [Giora et al.,](#)
109 [2015](#)) holds that, unless the sentence form bears a conventional or by-default
110 association with irony,³ utterance literal, compositional meaning will nec-

³ So far, evidence for such by-default ironic meanings, outside conventionally ironic constructions, is limited to negative statements of the form ‘X is not the most Y’ ([Giora et al., 2015](#)).

111 necessarily be activated first before being rejected in favor of a contextually
112 computed ironic interpretation. On different grounds, authors like [Sperber](#)
113 [& Wilson \(2002\)](#), who hold that any pragmatic processing involves context-
114 based inference of speaker’s intentions, would also have to predict that non-
115 contextual cues can supplement, but not replace context in irony compre-
116 hension.

117 Summing up, two related research questions clearly emerge from the cur-
118 rent state of the art: one about the reliability of non-contextual cues, and
119 the other about the relative processing roles of contextual and non-contextual
120 cues. In [Experiment 1](#) of this paper we assess the discrimination of ironic
121 prosody relative to neutral prosody, as well as to positive or negative literal
122 prosody; we also test, in the exact same way, the discriminability of ironic
123 facial expression. (While the discrimination of ironic prosody has been previ-
124 ously investigated, to the best of our knowledge no such evidence is available
125 for ironic facial expression.) In [Experiments 2 and 3](#) we assess how the same
126 prosody and facial cues, as well as contextual information impact irony com-
127 prehension, using an act-out task inspired by [Kowatch et al. \(2013\)](#). Our
128 Hypothesis 1 is that in a categorization task ironic prosody and ironic facial
129 expression should allow correct discrimination of ironic items. In line with the
130 model put forward by [Kissine \(2016\)](#), as well as with the Direct Access view
131 ([Gibbs, 2002](#)), we predict that in the act-out tasks of [Experiments 2 and 3](#)
132 the presence of salient – albeit potentially less reliable – non-contextual cues
133 should prompt interpreters to bypass costlier contextual processing. That
134 is, our Hypothesis 2 is that ironic prosody and facial expression are privi-
135 leged in irony comprehension at the expense of costlier, but more accurate

136 assessment of the utterance literal content relative to the context. Accord-
137 ingly, one should expect non-contextual cues to be associated with faster
138 responses; furthermore, if, as we predict, the processing of ironic prosody or
139 facial expression does not supplement context-based assessment of the com-
140 positional meaning, non-contextual cues should not entail any accuracy gain
141 relative to contextual incongruence.

142 Our [Hypothesis 2](#) may also be seen as one possible implementation of
143 the parallel-constraint-satisfaction model of irony interpretation ([Katz, 2005](#);
144 [Pexman, 2008](#)). As we already mentioned, this model predicts that contex-
145 tual and non-contextual cues are processed in parallel. If parallel processing
146 of all cues must be completed before the outputs are weighted and the fi-
147 nal interpretation reached, then, contrary to our predictions, the presence of
148 non-contextual cues along with contextual incongruence should lead to longer
149 reaction times. However, the parallel-constraint-satisfaction model could also
150 assign greater weight to non-contextual cues, in such a way that costlier pro-
151 cesses terminate before completion in the presence of salient ironic prosody
152 or facial expression. In that case, this model would be entirely consistent
153 with our predictions.

154 By contrast, the predictions generated by our [Hypothesis 2](#) are incom-
155 patible with accounts that posit obligatory processing of contextual cues in
156 irony derivation. According to Relevance theory ([Sperber & Wilson, 2002](#))
157 or the Graded Salience theory ([Giora, 2003](#)), ironic interpretation necessarily
158 involves the assessment of the utterance content against the background con-
159 text.⁴ If so, non-contextual cues would merely add a supplementary source of

⁴ Again, with the possible proviso concerning conventional or by-default ironic meanings

160 evidence for (or against) ironic interpretation, the integration of which should
161 lead to increased response times – especially if these cues are not entirely
162 congruent with context-based processing. Furthermore, if non-contextual
163 cues supplement obligatory context-based processing, accuracy levels should
164 increase when ironic prosody or facial expression combine with contextual
165 incongruence.

166 **Experimental stimuli**

167 All experiments reported in this paper use (part of) thirty-six videos (and
168 two practice trials videos) in French, in which two individuals discuss two
169 items placed on the table in front of them. Each video can be subdivided
170 in (*a*) a context segment, (*b*) a labeling and question segment, (*c*) a pause
171 segment and (*d*) the target utterance. First, the character (A) at the right
172 of the screen mentions her/his knowledge about the second character’s (B)
173 preferences regarding the two items placed on the table (e.g. ‘George, I
174 know that you like chemistry and that you really hate physics. But reading
175 a physics book could be interesting.’). This part contains contextual back-
176 ground information useful for detecting potential sarcasm. Second, A labels
177 the two items on the table to ensure participants could identify them (e.g.
178 ‘Here is a chemistry book and here is a physics book.’) and then asks B
179 if s/he wants one of the two items (e.g. ‘Would you like the physics book
180 as a gift, now?’). Third, a black screen appears and participants are asked
181 to press the space bar to hear B’s reply. Fourth, a video appears with B’s

(Giora et al., 2015); see footnote 3.

182 reply, viz. the target (e.g. ‘No, you know how much I hate physics!’). In
 183 Experiments 2 and 3 the video freezes until the participant selects the item
 184 she believes B really wants. Clicking the right mouse button corresponds to
 185 the object at the right of the screen and clicking the left mouse button to the
 186 object at the left of the screen. Time course of a video stimulus is illustrated
 187 in Figure 1.

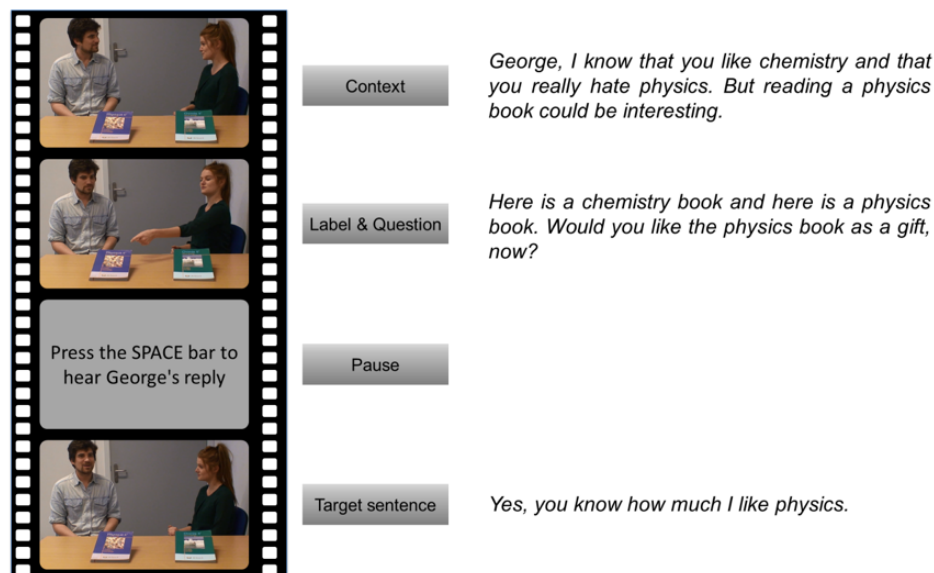


Figure 1: Time course of a full stimulus

188 The thirty-six videos are drawn on twelve scenarios (see Figure 2). There
 189 are three versions of each scenario based on the meaning of the target: an
 190 *Ironic*, a *Literal Yes* and a *Literal No* version (see Table 1 for a translated
 191 example of the three versions of a scenario). The meaning of the target was
 192 manipulated by modifying the contextual information (B’s preferences are
 193 congruent vs. incongruent with B’s reply) and the beginning of the target

194 (‘Yes, you know how much I like X’ vs. ‘No, you know how much I hate X’).
 195 Three professional actors formed three pairs (actor 1 and 2, actor 1 and 3,
 196 actor 2 and 3). Each actor performed in 12 videos, 6 as individual A and 6
 as individual B.

	LITERAL NO	IRONIC	LITERAL YES
<i>Context</i>	George, I know that you like drink- ing milk and that you really don’t like drinking tea for breakfast. But some kinds of tea are really nice.		George, I know that you like tea for breakfast and you have said this to me many times.
<i>Label</i> <i>Q</i> <i>Question</i>	George, here is a glass of milk and here is a cup of green tea. Would you like the cup of green tea with your break- fast, now?		
<i>Target</i>	No, you know how much I hate tea for breakfast!	Yes, you know how much I like tea for breakfast!	

Table 1: Example of three versions (Literal No, Ironiic and Literal Yes) of a scenario

197

198 Each video vignette was assigned to one of 6 conditions defined in func-
 199 tion of the presence or the absence of specific cues: Context only, Prosody
 200 only, Context & Prosody, Context & Facial expression, Prosody & Facial
 201 expression, and Context, Prosody & Facial expression. In conditions where
 202 contextual cues were not available, the context segment of the video was
 203 removed. In the conditions where the prosodic cues were not available, the
 204 actor was asked to utter the target sentence on a monotonous tone of voice.
 205 By contrast, when prosody cues were present, the actor was instructed to
 206 utter the target sentence with the corresponding prosody (literal positive for

207 Literal Yes items, literal negative for Literal No items and ironic for Ironic
 208 items). The same applies to facial expression cues. Each target is thus as-
 209 sociated with one of the four following prosody contours and one of four
 210 following facial expressions: *Ironic*, *Literal Yes*, *Literal No* and *Neutral*.

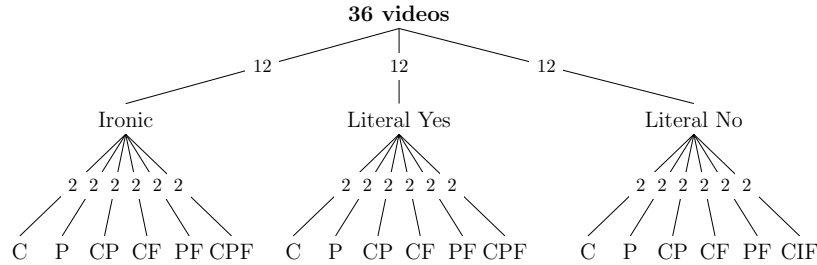


Figure 2: Assignment of the 36 experimental videos across conditions. Contextual cue (C), Prosody cue (P), Facial expression cue (F)

211 Recall that our overarching objective is to disentangle potentially differ-
 212 ential roles of contextual and non-contextual cues in irony processing. In
 213 order to do so, one should avoid using stimuli whose ironic character is in-
 214 herently difficult to grasp, as this would entail a markedly low accuracy on a
 215 sub-set of items. This is why, as in [Kowatch et al. \(2013\)](#), our ironic stimuli
 216 always consist of negative meanings associated with literally positive sen-
 217 tences (‘Yes’ sentences; see [Table 1](#)); utterances of the opposite valence —
 218 literally negative sentences with a positive ironic meaning — are much less
 219 canonical forms of irony, and have been found to be particularly difficult to
 220 grasp (e.g. [Kreuz & Link, 2002](#); [Climie & Pexman, 2008](#); [Filippova & Astington, 2010](#)). For a similar reason, no item is associated with Facial expression
 222 as the only cue towards the (non-)ironic meaning. A marked facial expres-
 223 sion with no context support and combined with a neutral prosody would be

224 too an unnatural and ambiguous cue, and would increase the risk of chance
225 performance.

226 It is possible that the most salient acoustic correlates of ironic prosody are
227 not intrinsic, but rather relative to the surrounding discourse (Bryant, 2010).
228 However, ironic prosody has also been reported to have inherent acoustic
229 correlates at the level of fundamental frequency (F0), intensity and delivery
230 rate (e.g. Rockwell, 2000; Bryant, 2010; Anolli et al., 2000; Løevenbruck et al.,
231 2013). For all target segments, F0 (in Hz, every 3 ms), intensity (in dB, every
232 11 ms) and syllable duration (in ms) were measured using Praat (Boersma &
233 Weenink, 2017). A linear regression implemented in the lmer package (Bates
234 et al., 2015) in R (R Core Team, 2016) reveals that only mean intensity
235 predicts the type of prosody ($F(3, 32) = 6.84$; $p = 0.001$; all other $ps > 0.3$).
236 Because the first word of the target is always a monosyllabic *yes* (*oui*) or
237 *no* (*non*), it made sense to assess whether ironic prosody is reflected in its
238 acoustic properties. For first syllables of target utterances, we found an effect
239 of prosody type on the mean intensity ($F(3, 32) = 4.472$, $p = 0.001$) and of
240 syllable length ($F(3, 32) = 8.586$, $p < 0.001$). As can be seen from Table 2,
241 the ironic prosody of our stimuli is associated with a significantly higher mean
242 intensity in the whole sentence and in the first syllable in comparison with
243 all other prosody types, and with significantly longer first syllable relative to
244 Neutral prosody.

245 In an attempt to objectify differences in facial expressions, five compo-
246 nents have been analyzed during the target sentence: eyes, mouth, eyebrows,
247 head and upper body. If there was at least one movement in a component
248 (e.g., raising eyebrows), it was scored as 1; if the component remained still,

	Mean intensity (sentence)	Mean intensity (first syllable)	Length (first syllable)
Intercept (Ironic)	73.48*** (0.79)	71.54*** (0.96)	553.50*** (41.34)
Literal Yes	-3.35** (1.11)	-4.22** (1.36)	-106.37 (58.47)
Literal No	-3.68** (1.11)	-0.31 (1.36)	-102.50 (58.47)
Neutral	-4.43*** (1.01)	-2.64* (1.24)	-261.33*** (53.38)
R ²	0.39	0.30	0.45
Adj. R ²	0.33	0.23	0.39
Num. obs.	36	36	36
F statistic	6.84	4.47	8.59

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 2: Coefficients (and standard errors) of linear regressions of mean intensity (whole sentence and first syllable) and of syllable length (first syllable) on prosody Type

249 it was scored as 0. A total facial expression score (ranging from 0 to 5) for
250 an item is then the sum of the five component scores. Actors have been
251 consistent in their way to display emotions. For neutral facial expressions
252 ($n = 18$), they all kept still ($m = 0.28, sd = 0.57$). For marked facial ex-
253 pressions (ironic or literal, $n = 18$), actors were all using combinations of
254 different movements ($m = 2.94, sd = 1.26$). An ordinal regression imple-
255 mented in the `clm` function from the `ordinal` package (Christensen, 2015) in
256 R (R Core Team, 2016) reveals a significant effect of video category (Marked
257 vs. Neutral facial expression; $z = 34.98, p < 0.001$) with a higher number
258 of component movements in the Marked facial expression condition, but no
259 effect of Type (Ironic vs. Literal Yes vs. Literal No; $p > 0.19$).

260 Further qualitative inspection reveals a difference between Marked facial

expressions across Types. In the Literal Yes condition, actors look enthusiastic or happy: they display sincere static smile (mouth and eyes), as well as little and brief eyebrows upward movements. In the Literal No condition, actors look upset or angry: they visibly accentuate plosive consonants, shrug and slightly project their torsos, display accentuated and long eyebrow raising and frowning, produce head negation movements at the syllable rhythm, as well as increased blinking, half-closure and a wide-open eyes. In the Irony condition actors look sarcastic: they produce many wide eyebrow upward movements, eyebrows are also often arched, they display huge false frozen smiles ending in a cold expression, sway their body, and produce many repeated wide and slow head movements.

Experiment 1

The aim of Experiment 1 is to assess whether prosody or facial expressions can be correctly discriminated as ironic against a sincere or neutral counterpart. To this end, we isolated these cues from target segments of our material, and ran a first experiment where participants had to rate these isolated cues from sincere to ironic on a 7-point Likert scale.

Participants

One hundred thirty-nine volunteers participated in this first experiment. Data from 12 participants were discarded from the analyses because they were not French native speakers ($N=9$) or due to a problem with the network connection during the task ($N=3$). The remaining 127 volunteers (63 males) ranged in age from 16 to 53 years ($m = 26.68$, $sd = 7.50$).

284 Procedure

285 To investigate the discrimination of ironic prosody and facial expression,
286 we isolated the targets of the 36 videos described in the [previous section](#),
287 and next, split them into a stand-alone audio file and a video file with muted
288 sound. The resulting audio and video files all belong to one of the following
289 four types: *Literal Yes*, *Literal No*, *Neutral* and *Ironic*. The 36 audio files
290 and 36 video files were presented in the same randomised order across par-
291 ticipants, using the online survey application [LimeSurvey](#), with implemented
292 audio and video players. The experiment was composed of two parts: the
293 scoring of the facial expression and the scoring of the prosody. One group of
294 participants (Group *Prosody-Facial Expression*: $N=66$, 43 males, age 16-49
295 years, $m = 26.06$, $sd = 7.04$) scored first the prosody and then the facial
296 expression, while the second group performed the task in the opposite or-
297 der (Group *Facial Expression-Prosody*: $N= 61$, 20 males, age 19-53 years,
298 $m = 27.34$, $sd = 7.98$). In the prosody part, participants were asked to listen
299 to each sound excerpt and to rate the speaker’s tone of voice on a 7-point
300 Likert scale ranging from (1)-‘completely sincere’ to (7)-‘completely ironic’.
301 They were instructed to focus on the prosody independently from the con-
302 tent of the sentence. In the facial expression part, participants were asked
303 to watch each video and to rate the (left side of the screen) speaker’s facial
304 expression on the same 7-point Likert scale. We pointed out that the sound
305 of the videos was removed to allow them to focus on the speaker’s facial
306 expression.

307 *Results*

308 Participants' ratings of prosody and facial expression were analysed with
309 cumulative link mixed models, using the `clmm` function from the `ordinal`
310 package in R (Christensen, 2015). Here and in Experiments 2 and 3, the
311 significance of the fixed effects was assessed by performing likelihood ratio
312 tests in which a model containing the fixed effect is compared to another
313 model without it, but with an otherwise identical structure (Baayen et al.,
314 2008). Post-hoc comparisons of least square-means, with Tukey adjustment
315 for multiple comparisons and Satterthwaite method for estimating degrees of
316 freedom, were performed using the `lsmeans` package (Lenth, 2016).

317 *Prosody (audio data)*. Figure 3 displays irony rating on a (1-7) Likert scale
318 per type of prosody. Cumulative link multilevel logit regressions with by-
319 subject random intercepts and random slopes for the Type factor (Ironic
320 vs. Literal Yes vs. Literal No vs. Neutral) revealed a significant effect
321 of Type ($\chi^2(3) = 233.18$, $p < 0.001$), but not of Group (Prosody-Facial
322 Expression vs. Facial Expression-Prosody) ($p = 0.8$). The model, displayed
323 in Table 3, shows that Ironic prosody leads to significantly higher irony scores
324 than all other types of prosody. Since, however, other levels of Type do not
325 seem equivalent (see Figure 3), we conducted post-hoc comparisons, which
326 confirmed that Literal No type was rated as significantly less ironic than
327 Literal Yes ($z = -16.48$, $p < 0.001$) and Neutral ($z = -17.37$, $p < 0.001$),
328 while there was no difference between Literal Yes and Neutral types ($p =$
329 0.92).

330 *Facial expression (Video data)*. Figure 4 displays irony rating on a (1-7)
331 Likert scale per type of facial expression. Cumulative link multilevel logit

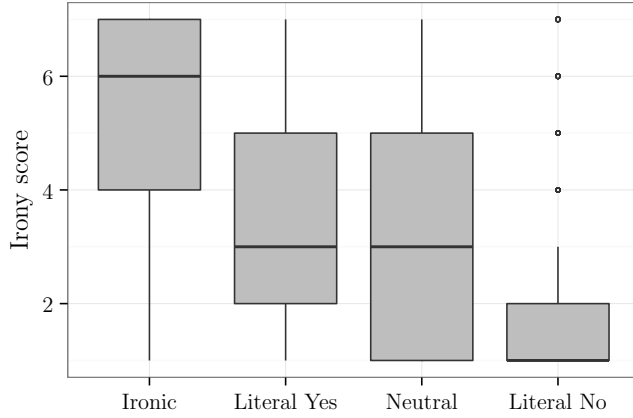


Figure 3: Tukey box-plots for ratings of audio-files per Prosody type

332 regressions with by-subject random intercepts and random slopes for the
 333 Type factor (Ironic vs. Literal Yes vs. Neutral vs. Literal No) revealed a
 334 significant effect of Type ($\chi^2(3) = 176.58$, $p < 0.001$), but not of Group
 335 (Prosody-Facial Expression vs. Facial Expression-Prosody; $p = 0.38$). As
 336 shown in Table 4, Ironic facial expression prompts significantly higher irony
 337 scores relative to other types of facial expression. Again, Figure 4 suggests
 338 that not all non-ironic levels are equivalent; post-hoc comparisons show that
 339 Literal Yes type is judged more ironic than Neutral ($z = 7.9$, $p < 0.001$) and
 340 Literal No ($z = 11.81$, $p < 0.001$), and that Neutral type is judged more
 341 ironic than Literal No ($z = 5.17$, $p < 0.001$).

342 Discussion

343 Experiment 1 confirms that in a rating task that explicitly opposes ironic
 344 to literal stimuli, ironic prosody and facial expression can be correctly dis-
 345 criminated against literal (positive or negative) or neutral prosody and facial

Irony ratings of prosody	
Literal Yes	−1.38 (0.09)***
Neutral	−1.62 (0.09)***
Literal No	−3.61 (0.15)***
1—2	−2.78 (0.08)***
2—3	−1.84 (0.07)***
3—4	−1.32 (0.07)***
4—5	−0.76 (0.07)***
5—6	−0.16 (0.06)*
6—7	0.68 (0.06)***
Num. obs.	5040

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 3: Coefficients (and standard errors) of fixed effects of the cumulative link multilevel logit regressions of irony ratings on Prosody Type on irony ratings (Ironic prosody is the reference level). The three first lines of the table report the coefficients for the predictors included in the CLMM. Lines 4 to 9 report the coefficients for each transition of the 7-point Likert scale.

346 expression. However, using a gradual Likert scale instead of a forced ‘ironic
 347 vs. literal’ binary choice (unlike, e.g. [Bryant & Fox Tree, 2002](#); [Voyer et al.,](#)
 348 [2016](#)) allows a finer-grained insight into the identification of prosody and
 349 facial expression as ironic or not. Of course, ironic prosody and facial ex-
 350 pression led to unambiguously higher scores on an irony scale. (In the case of
 351 ironic prosody, such perceptual judgements thus reflect the acoustic salience
 352 of the stimuli intensity.) Recall, however, that our non-ironic audio and video
 353 stimuli fall into three different types: positive prosody/expression, negative
 354 prosody/expression and neutral prosody/expression. If ironic cues were com-
 355 pletely unambiguous, one should expect the remaining three types to receive
 356 the same ironic scores. And yet, literal positive prosody contours and facial

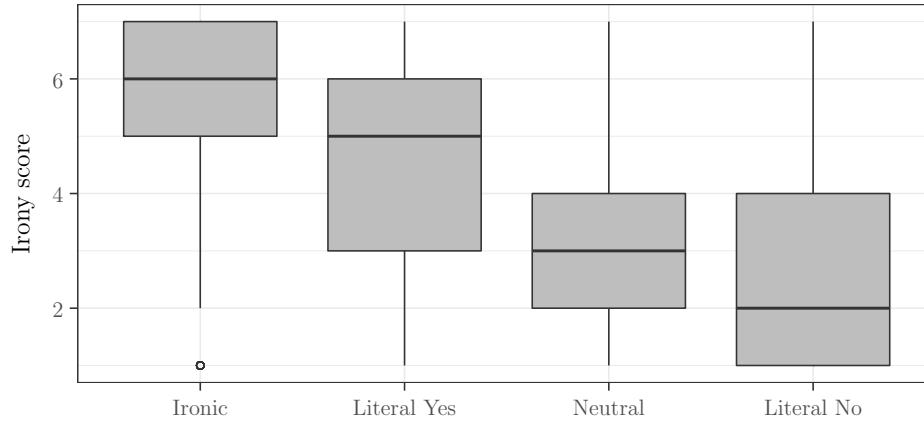


Figure 4: Tukey box-plots for ratings of video-files per facial expression type

expressions were judged more ironic than their literal negative counterparts. In other words, positive or neutral prosody and facial expression are more ambiguous as to the ironic vs. sincere meaning of an utterance. Furthermore, while ironic cues were accurately discriminated against the other ones as more ironic, the distinction could have been artificially boosted up by the fact that participants' task is reduced to merely rating the ironic dimension of various stimuli. In real life, however, interpreters have to decide on speaker's intentions rather than classify the utterance along an ironic-literal continuum. It is therefore likely that, in such settings, the actual reliability of non-contextual cues to irony is considerably lower than what the results of Experiment 1 might suggest.

Experiment 2

The second experiment uses an act-out irony comprehension task in order to compare the impact of ironic prosody and facial expression relative to

Irony ratings of facial expressions	
Literal Yes	-1.42 (0.14)***
Neutral	-2.64 (0.15)***
Literal No	-3.28 (0.18)***
1—2	-4.21 (0.13)***
2—3	-3.10 (0.13)***
3—4	-2.42 (0.12)***
4—5	-1.58 (0.12)***
5—6	-0.71 (0.12)***
6—7	0.28 (0.12)*
Num. obs.	5292

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 4: Coefficients (and standard errors) of fixed effects of the cumulative link multilevel logit regressions of irony ratings on Facial expression Type (Ironic facial expression is the reference level). The three first lines of the table report the coefficients for the predictors included in the CLMM. Lines 4 to 9 report the coefficients for each transition of the 7-point Likert scale.

that of contextual incongruence. First, we expect that, in spite of being correctly discriminated in [Experiment 1](#) ironic prosody and facial expression, should not improve accuracy in irony comprehension relative to contextual incongruence. Second, in spite of not being associated with an accuracy gain relative to context, we expect these non-contextual cues to lead to shorter response times.

Participants

Fifty-six students gave their written informed consent to participate in this study approved by the Faculty Ethics Committee at the Université libre de Bruxelles. Participants were recruited according to the following criteria: native French speakers, normal or corrected-to-normal vision, no hearing dif-

382 faculties, no history of neurological disorders. Ten participants were excluded
383 from statistical analyses because they were not native speakers of French.
384 Two other participants were excluded because they reported a history of
385 attention-deficit/hyperactivity disorder, a neurodevelopmental disorder as-
386 sociated with pragmatic deficits, including difficulties understanding irony
387 (e.g. [Caillies et al., 2014](#); [Staikova et al., 2013](#); [Bignell & Cain, 2007](#)). The
388 age of the 44 remaining participants (15 males) ranged between 18 and 26
389 years ($m = 20.43$; $sd = 1.47$).

390 *Procedure*

391 The task was run in 64-bit Windows 7 using Tobii StudioTM 3.2.1 soft-
392 ware, which controlled the stimuli presentation in a random order and recorded
393 participant’s response and reaction times. A Tobii pro X2-60(Hz) screen-
394 based eye-tracker device (Tobii Technology, Inc. Stockholm, Sweden) was
395 used to record participants’ eye movements during the target sentence. A
396 five-point calibration procedure designed by Tobii Studio was used before the
397 irony task.

398 Each participant was seated at a distance of ± 60 cm in front of a 16.5-inch
399 monitor (resolution: 1920 x 1080 pixels) wearing headphones. The stimuli
400 were presented at a comfortable sound pressure level (65dB \pm 5dB). Follow-
401 ing eye-tracker calibration, participants were presented with the following
402 instructions on screen:

403 In each trial of this task you will watch videos with short conver-
404 sations between two individuals. One person will ask a second
405 person questions about two items on a table. After each ques-
406 tion, you will watch a video with the second person’s reply. Listen

carefully to what the first person says and to the second person's
reply! At the end of the second person's answer you will have
to give him/her the item you believe he/she really wants. You
should press the left mouse button if you think that he/she re-
ally wants the item at the left or the right mouse button if you
believe that he/she really wants the item at the right. Next trial
will start automatically after your answer.

Participants first completed two training trials (one Literal Yes and one Lit-
eral No item) before the experimental phase began.

Results

In order to assess the relative impact of context, prosody and facial ex-
pression, each stimulus was associated with a binomial variable Context,
Prosody and Facial expression, depending on which cue(s) were associated
with the target. Note that our Sincere No items were unambiguously literal.
It is possible, then, that participants realize, in the course of the experiment,
that any answer starting with 'No' would lead to a non-ironic interpretation.
In order to assess this possibility, we examined the effect of the linear Order
of the stimuli.

Accuracy. A correct interpretation of a target corresponds to a trial where
the participant accurately selects the object the second character (B) in the
video really wants (see [the description of the stimuli](#)). For Literal Yes items,
the correct choice was the object named in the target (e.g. 'Yes, you know
how much I like physics!'), whereas in Literal No (e.g. 'No, you know how
much I hate physics!') and Ironic items (e.g. 'Yes, you know how much I

like physics!'), it was the other object displayed in the video. As can be seen from Figure 5, while accuracy rate is generally high, it is lower for Ironi

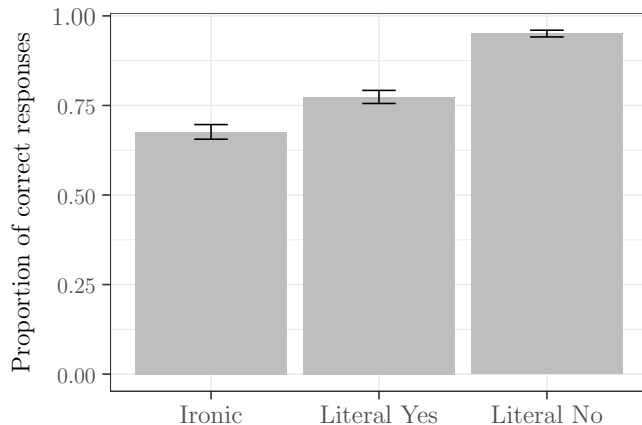


Figure 5: Proportions of correct responses by utterance Type (vertical bars represent standard errors)

Binomial logistic multilevel models, with the by-participant random intercepts were implemented using the `glmer` function of the `lme4` package (Bates et al., 2015). They revealed a significant effect of Type (Ironi vs. Literal Yes vs. Literal No; $\chi^2(2) = 149.62$, $p < 0.001$), as well as of Context ($\chi^2(1) = 8.75$, $p = 0.003$) and of Facial Expression ($\chi^2(1) = 814.85$, $p = 0.001$); there was no effect of Prosody ($p = 0.14$) and of Order ($p = 0.21$). Interactions between Type and Context ($\chi^2(2) = 10.4$, $p < 0.007$) and Type and Face ($\chi^2(3) = 77.85$, $p < 0.001$) were also significant. The model in Table 5 shows that Ironi targets elicit less correct responses than Literal No ones. The presence of Context strongly increases the accuracy on Ironi items. As for Facial expression, it has a detrimental effect on accuracy of

445 Ironic items; on the contrary, it increases the accuracy of Literal Yes relative
446 to Ironic items.

	Accuracy
Intercept (Ironic)	0.45 (0.20)*
Literal No	2.68 (0.46)***
Literal Yes	0.05 (0.28)
Context	
Ironic X Context	0.86 (0.20)***
Literal No X Context	-0.33 (0.45)
Literal Yes X Context	0.07 (0.24)
Facial expression	
Ironic X Facial expression	-0.44 (0.20)*
Literal No X Facial expression	0.32 (0.41)
Literal Yes X Facial expression	1.98 (0.26)***
Num. obs.	1578

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 5: Coefficients (and standard errors) of fixed effects for the multilevel logistical regression of correct responses on target Type, Type X Context & Type X Facial expression

447 *Reaction times.* In each target, the speaker’s preference was entirely deter-
448 mined once the word was uttered referring to the object s/he wanted or pre-
449 tended to want (‘Yes, you know how much I like **physics!**’). Reaction times
450 were recorded from the onset of the word referring to an object in the target
451 until participant’s response. A negative reaction time means that participant
452 responded before the onset of the target word. Boxplots in [Figure 6](#) suggest
453 longer reaction times for Ironic targets. Linear multilevel regression models,
454 with by-participant random intercepts and random slopes for the Type factor,
455 were implemented with the `lmer` function of the `lme4` package ([Bates et al.,](#)
456 [2015](#)). They revealed a significant effect of Type ($\chi^2(2) = 31.36$, $p < 0.001$),

457 Prosody ($\chi^2(1) = 11.85$, $p < 0.001$), and Facial Expression ($\chi^2(1) = 5.3$,
 458 $p = 0.021$), but not of Context ($p = 0.48$). Interactions between Type and
 459 Prosody ($\chi^2(1) = 8.45$, $p = 0.15$) and between Type and Facial expression
 460 ($\chi^2(2) = 21.166$, $p < 0.001$) were also significant. Finally, there was also
 461 an effect of Order ($\chi^2(1) = 146.87$, $p < 0.001$), but no interaction between
 462 Order and Type ($p = 0.61$). The model, displayed in [Table 6](#) reveals that
 463 responding to Ironic targets takes longer than for the other two types and
 464 that reaction times decrease along experimental trials. Both Prosody and
 465 Facial Expression decrease reaction times for Ironic items. (Prosody also de-
 466 creases the processing of Literal Yes, but increases that of Literal No relative
 467 to Ironic items.)

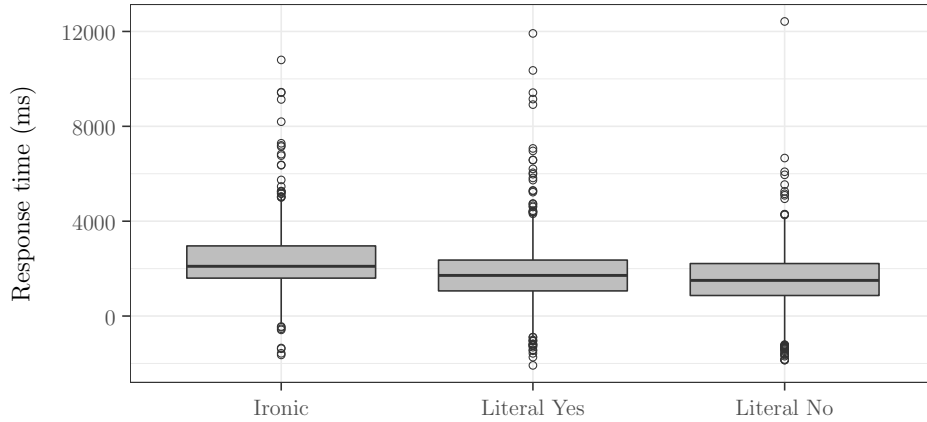


Figure 6: Tukey box-plots for reaction time per target type

468 *Eye-Tracking data.* We identified three areas of interest (AOI) for the target
 469 sentence segment using Tobii Studio software (version 3.2.1): the speaker's
 470 face (eyes plus lips regions), the correct and the incorrect objects (see [Figure](#)

	Reaction time
Intercept (Ironic)	3576.93 (220.77)***
Literal No	-1442.15 (174.05)***
Literal Yes	-754.39 (192.42)***
Order	-38.62 (3.07)***
Prosody	
Ironic X Prosody	-564.82 (121.77)***
Literal No X Prosody	216.16 (104.35)*
Literal Yes X Prosody	-321.43 (118.52)**
Facial expression	
Ironic X Facial expression	-457.33 (117.80)***
Literal No X Facial expression	-143.85 (98.26)
Literal Yes X Facial expression	14.45 (112.48)
Num. groups: Subject	44

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 6: Coefficients (and standard errors) of fixed effects of the multilevel linear regression of reaction time on target Type, Half , Type X Prosody & Type X Facial expression

7). The position of the AOI was manually adapted to the movements of the two actors in real time. Eye movement data for the target sentence segment was exported from Tobii using the I-VT fixation filter in the default setting. From the onset of the word referring to the object in the target, and for each AOI, we calculated the total fixation duration (i.e. the sum of the duration for all fixations within an AOI) and the fixation count (i.e. the number of times the participant fixates on an AOI). The fixation count and the total fixation duration were normalised according to participants' reaction times. For instance, number of fixations for the AOI 'correct object' = ([number of fixations on the correct object / time between the beginning of the target word until participant's response] * 1000). We also coded whether participants' first three fixations went to the correct object, both from the

483 onset of the target utterance and from the onset of the word referring to the
484 object in the target.



Figure 7: Areas of interest (AOIs) for the target sentence segment: the speaker's face (eyes plus lips regions), the correct and incorrect objects.

485 A linear mixed model on proportion of fixation durations with random
486 by-participant intercepts revealed a significant effect of AOI ($\chi^2(2) = 112.77$,
487 $p < 0.001$), as well as a significant AOI X Type interaction ($\chi^2(6) = 24.87$,
488 $p < 0.001$). There was no interaction between AOI and Context, AOI and
489 Prosody, and AOI and Facial expression (all $ps > 0.14$). A linear mixed
490 model on proportion of fixation counts with random by-participant intercepts
491 also revealed a significant effect of AOI ($\chi^2(2) = 69.12$, $p < 0.001$) and a
492 significant AOI and Type ($\chi^2(6) = 23.76$, $p < 0.001$). Additionally, there
493 was a weak interaction between AOI and Context ($\chi^2(3) = 8.69$, $p = 0.034$),

494 but no interaction between AOI and Prosody ($p = 0.88$), between AOI and
 495 Facial expression ($p = 0.086$), and between AOI, Type and Context ($p = 0.4$).
 496 As can be seen from the model summaries in Table 7, the most relevant result
 497 – which is also fairly consistent with accuracy data – from fixation duration
 498 and counts is that Literal No targets attracted longer and more numerous
 499 fixations to the correct object than Irony ones.

	Proportions of fixation durations	Proportions of fixation counts
Intercept(Correct Object)	4.96 (2.64)	9.19 (3.04)**
Incorrect Object	-0.51 (2.63)	-3.34 (3.42)
Speaker's face	11.37 (2.28)***	8.94 (2.96)**
Type		
Correct Object X Literal Yes	2.70 (2.64)	2.94 (2.64)
Incorrect Object X Literal Yes	-1.23 (2.64)	-1.19 (2.64)
Speaker's face X Literal Yes	2.06 (1.87)	1.55 (1.87)
Correct Object X Literal No	11.76 (2.66)***	11.92 (2.66)***
Incorrect Object X Literal No	1.21 (2.66)	1.43 (2.66)
Speaker's face X Literal No	3.03 (1.88)	1.73 (1.88)
Correct Object X Context		-3.85 (2.30)
Incorrect Object X Context		-0.46 (2.30)
Speaker's face X Context		-3.93 (1.63)*
Num. obs.	5609	5608

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 7: Coefficients (and standard errors) of fixed effects of the multilevel linear regression of proportions of fixation durations and fixation counts on AOI, AOI X target Type, & AOI X Context

500 Turning to the first three fixations on the correct object from the be-
 501 ginning of the target utterance, binomial multilevel models, with the by-
 502 participant random intercepts revealed an effect of Type ($\chi^2(2) = 35.56$,

503 $p < 0.001$), but no effect of Context, Prosody or Facial expression (all
 504 $ps > 0.06$). There was also an effect of the Fixation Number (first, sec-
 505 ond or third; $\chi^2(1) = 32.93$, $p < 0.001$), as well as interaction between Type
 506 and Fixation Number ($\chi^2(3) = 36.16$, $p < 0.001$).

507 For the first three fixations on the correct object, computed from the
 508 beginning of the word referring to the object, binomial multilevel models,
 509 with the by-participant random intercepts also revealed an effect of Type
 510 ($\chi^2(2) = 19.17$, $p < 0.001$) on the fixation on the correct object, but no
 511 effect of Context, Prosody or Facial expression (all $ps > 0.21$). Here too,
 512 there was an effect of Fixation Number ($\chi^2(1) = 123.02$, $p < 0.001$), as well as
 513 interaction between Type and Fixation Number ($\chi^2(3) = 125.03$, $p < 0.001$).
 514 As can be seen from [Table 8](#), from the start of the target utterance, Literal
 515 Yes – but not Literal No – items attract more anticipatory fixations to the
 516 correct object than Ironic. Towards the end of the target utterance, however,
 517 Literal No items – but not Literal Yes – are more likely to attract anticipatory
 518 looks towards the correct object than Ironic items. For both measures (i.e.
 519 target utterance and object mention), the probability to look at the correct
 520 object increases from the first to the third fixation for all item types.

521 *Discussion*

522 Correct identification of the speaker’s goals is significantly lower for ironic
 523 utterances; in that respect, our results parallel findings by [Kowatch et al.](#)
 524 ([2013](#)), who used analogous utterance Types. Equally consistent with the
 525 literature (e.g. [Kreuz & Roberts, 1995](#); [Gerrig & Goldvarg, 2000](#)) is our result
 526 that incongruence with the preceding context increases the correct interpre-
 527 tation of ironic utterances. Interestingly, ironic prosody does not facilitate

	First fixations (target utterance)	First fixations (object mention)
Intercept (Ironic)	-5.30 (0.54)***	-3.90 (0.32)***
Literal Yes	1.42 (0.59)*	0.39 (0.41)
Literal No	0.03 (0.67)	0.78 (0.38)*
Fixation number		
Ironic X Fixation number	0.58 (0.22)**	0.80 (0.13)***
Literal Yes X Fixation number	0.45 (0.14)***	0.72 (0.12)***
Literal No X Fixation number	0.76 (0.18)***	0.71 (0.11)***
Num. obs.	3905	3402
Num. groups: Subject	43	43

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 8: Fixed effects of the multilevel logistic regression of first fixations on the correct object (beginning from start the target utterance and from the start of the word referring to an object) target Type & Type X Fixation Number (first, second or third).

528 interpretation, and ironic facial expression actually hampers it, confirming
529 that non-contextual cues for irony are not very reliable in a comprehension
530 task.

531 We also found that ironic items elicit slower reaction times relative to
532 the literal ones. Slower processing of ironic items, to a certain extent at
533 least, is probably linked to the contextual assessment and rejection of the
534 compositional, literal meanings (Giora, 2003). However, Experiment 2 also
535 strongly suggests that context-based processing of irony may be aborted in
536 the presence of a distinctive prosody and/or facial expression. A striking re-
537 sult, which is consistent with our predictions, is that both ironic prosody and
538 ironic facial cues dramatically decrease reaction times. Together, accuracy
539 and reaction times results indicate that non-contextual cues entail a trade-off
540 in irony interpretation. On the one hand, they are less reliable than context,

541 but, on the other hand, their presence prompts a faster processing. One
542 explanation of this effect, in line with the model proposed by [Kissine \(2016\)](#),
543 is that, in the presence of distinctive prosody and/or facial expression par-
544 ticipants by-pass contextual interpretation of the literal meaning. Another,
545 very similar interpretation is that both contextual and non-contextual cues
546 are processed in parallel, but that the latter lead to faster decision, thus
547 terminating the former. This finding is coherent with the parallel-constraint-
548 satisfaction account ([Katz, 2005](#); [Pexman, 2008](#)), but allows to go a step
549 further in evidencing the relative weight of the different cues in irony com-
550 prehension. Note that, unlike us, [Kowatch et al. \(2013\)](#) found no difference
551 in reaction times between ironic and literal items. Recall, however, that their
552 stimuli were all associated with ironic prosody and no contextual cues. To
553 the extent that prosody prompts faster (but less accurate) processing, this
554 feature may explain the difference between their and our results.

555 Another clear-cut result of Experiment 2 is the advantage in processing
556 for literal ('No') negative sentences. These items led to strikingly higher accu-
557 racy rates and faster response times. Likewise, these items were associated
558 with longer and more numerous fixations on the correct object, reflecting
559 lesser hesitation as to the response. Further evidence for the advantage of
560 Literal No items comes from first fixations. At the beginning of the target
561 utterance, more looks go to the correct object in literal positive items, which
562 is certainly due to the spill-over from the mention of the correct object in
563 the preceding question (see [Figure 1](#) and [Table 1](#)). However, by the time
564 the object is mentioned in the target, literal negative utterances trigger more
565 anticipatory looks towards the correct object. One reason why Literal No

566 items stand apart could be that in our design all ironic items were associ-
567 ated with positive ('Yes') sentences. However, neither accuracy nor reaction
568 times for negative sentences change across the experiment, as revealed by
569 the absence of the interaction between Type and Order. There is another
570 reason why participants implicitly grasped the unambiguously literal nature
571 of Literal No items. Irony is usually associated with negatively oriented read-
572 ings of positive literal counterparts, whereas the opposite, ironic positive /
573 literal negative valence is highly atypical (Kreuz & Link, 2002). One may
574 surmise, then, that interpreters' language experience makes them sensitive to
575 irony's negative valence. If so, our participants could rapidly associate nega-
576 tive sentences with a non-ironic interpretation, without necessarily assessing
577 the literal content relative to the context, or, for that matter, processing
578 any other cue. In line with this idea, even though in Experiment 1 negative
579 prosody was the most clearly distinguished from ironic, in Experiment 2 it
580 tended to slow down reaction times. That is, information provided by literal
581 negative prosody is made redundant by the negative valence of the sentence.

582 Returning to our main research questions, Experiment 2 strongly suggests
583 that, as hypothesised in the Introduction, ironic prosody and facial expression
584 are less reliable cues for irony comprehension than contextual incongruence,
585 but that they entail an accuracy-processing speed trade-off. However, three
586 methodological choices we made could be taken to somehow mitigate these
587 results. First, recall that we decided not to include a condition with Facial
588 expression, but no Prosody and Context. While the results of Experiment 2
589 indicate that ironic Facial expression is not a fully reliable cue for the inter-
590 pretation of ironic utterances, this cue was always associated with at least

591 another one. Second, our decision not to include Ironic No items (viz. ironic
592 compliments), albeit fully justified from a methodological point of view (see
593 above), entails an unbalanced experimental design, with more positive (Ironic
594 and Literal Yes) than negative (Literal No) items. Third, we compared the
595 reliability of a prosodic or contextual cue delivered alone or in combination
596 with one or two other cues, but at no time these conditions were compared
597 to a complete absence of cues. Adding a control condition without any cues
598 would allow to appreciate more closely the reliability of isolated cues. Even
599 though none of these three features is likely to impact the differential pro-
600 cessing roles of contextual and non-contextual cues uncovered in Experiment
601 2, we seek to determine, in Experiment 3, whether these effects are robust
602 enough to show up in a fully balanced design.

603 **Experiment 3**

604 In Experiment 3 we seek to replicate the effects uncovered in Experiment
605 2 using a perfectly balanced, between-subject design, while keeping exactly
606 the same target sentences segments as in Experiments 1 and 2. In order to
607 rule out any potential bias due to the presence of literal negative items, we
608 kept only the Literal Yes and Ironic items of Experiment 2.⁵ Using these

⁵ In theory, we could have balanced our design by adding negative ironic items. To begin with, such a design would have considerably increased the task duration, and hence the risk of biases due to cognitive fatigue. More importantly, and as already discussed above, negative ironic items are highly atypical (Kreuz & Link, 2002), and poorly comprehended even in discrimination tasks, which arguably are easier than our act-out paradigm (Climie & Pexman, 2008; Filippova & Astington, 2010). Interpretation of this less common type

609 items, we created two sets of stimuli – Context vs. No Context – presented
610 to two different groups of participants. Stimuli presented to the Context
611 Group included all possible combinations of Context with non-contextual
612 cues: Context; Context and Facial expression; Context and Prosody; and
613 Context, Prosody and Facial expression. Stimuli presented to the No Con-
614 text Group pooled all combinations of non-contextual cues in the absence of
615 Context: No cue; Prosody; Facial expression; Prosody and Facial expression.
616 In this way, Experiment 3 provides a balanced design suited to isolate the im-
617 pact of all three cues, and includes a condition with only a facial-expression
618 cue and a control condition without any cues.

619 This between-subject design also allows to further test our Hypothesis
620 2. In line with Experiment 2, we expect an overall effect of Group (Context
621 vs. No Context) on accuracy: in the absence of contextual cues, partici-
622 pants should be more error prone in gauging the speaker’s ironic intent. If,
623 as we hypothesise, processing of non-contextual cues is privileged at the ex-
624 pense of context-based assessment of the utterance content, the presence of
625 ironic prosody and/or facial expression should lead to comparable accuracy in
626 both Context and No Context groups. Furthermore, we also predict contex-
627 tual processing not to be completed in the presence of these non-contextual
628 cues. Accordingly, the presence of the non-contextual cues lead to shorter
629 response times in both groups, which would indicate, in the Context group,
630 that contextual processing has been aborted. If, by contrast, non-contextual
631 cues supplement full processing of contextual cues, one should expect ironic

of irony is thus a topic orthogonal to the relative roles of contextual and non-contextual cues, and clearly falls out of the scope of this paper.

632 prosody and facial expression to increase accuracy in the Context group, and
633 to lead to longer response times.

634 *Participants*

635 Forty-seven undergraduate students, none of whom took part in Experi-
636 ments 1 and 2, participated for monetary reward in Experiment 3. Inclusion
637 criteria were similar to Experiment 2. One participant was excluded from sta-
638 tistical analyses for technical reasons. The Context group (N=23) consisted
639 of 13 women and 10 men between 19 and 29 years ($m = 22.70$; $sd = 2.60$),
640 and the No Context group consisted of 15 women and 8 men between 20 and
641 28 years ($m = 22.83$; $sd = 2.50$).

642 *Procedure*

643 Two groups of sixteen videos from the previous set of videos were used
644 to form the Context and No Context group. Target sentences in the Context
645 group were always associated with a Contextual cue and were subdivided
646 in 4 categories depending on the presence (+) or the absence (-) of prosody
647 and/or facial expression cues: Context only (C+P-F-), Context & Prosody
648 (C+P+F-), Context & Facial expression (C+P-F+), Context, Prosody &
649 Facial expression (C+P+F+) conditions. In the No Context group, the tar-
650 get sentence was never associated with a Contextual cue; stimuli were also
651 subdivided in 4 categories depending on the presence (+) or absence (-) of
652 prosody and facial expression cues: No cues (C-P-F-), Prosody only (C-P+F-
653), Facial expression only (C-P-F+), Prosody & Facial expression (C-P+F+)
654 conditions. Each category is composed of 2 Ironic and 2 Literal Yes items
655 (see [Table 9](#)).

Group	Context			No Context		
	Context	Prosody	Facial expression	Context	Prosody	Facial expression
	+	-	-	-	-	-
	+	+	-	-	+	-
	+	-	+	-	-	+
	+	+	+	-	+	+

Table 9: Design of Experiment 3

656 To obtain the items in the No cues (C-P-F-) and in the Facial expres-
657 sion only (C-P-F+) conditions, we used the same videos as in the Context
658 only and Context & Facial expression conditions from the Context group,
659 removing the context segment from the videos.

660 *Results*

661 In order to uncover the roles of ironic prosody and facial expression, as
662 in [Experiment 2](#), we associated each item with binomial Prosody and Facial
663 Expression factor, depending on which cue(s) were associated with the target.

664 *Accuracy.* [Figure 8](#) displays the proportions of correct responses by Group,
665 Type and non-contextual cue. Accuracy was analyzed building hierarchical
666 binomial logistic multilevel models, with the by-participant random inter-
667 cepts, using the `glmer` function of the `lme4` package ([Bates et al., 2015](#)). As
668 predicted, there was a significant effect of Group (Context vs. No Con-
669 text; $\chi^2(1) = 7.94$, $p < 0.005$). There was also an effect of Type (Literal
670 vs. Ironic; $\chi^2(1) = 5.78$, $p < 0.001$), but no Group X Type interaction
671 ($p = 0.42$). There was no effect of Prosody ($p = 0.17$), but an effect of

672 Facial expression ($\chi^2(1) = 36.2, p < 0.001$). The Type X Facial expression
673 interaction was significant ($\chi^2(1) = 8.32, p < 0.004$), but not the Group X
674 Facial expression ($p = 0.49$). That is, ironic prosody and facial expression
675 have a comparable effect in both groups, indicating that their presence does
676 not have a cumulative effect on accuracy in the Context group.

677 In order to assess further our predictions, we conducted post-hoc com-
678 parisons of least square-means on the final model. As predicted, overall
679 accuracy is significantly lower in the No Context group ($\beta = -0.53, se =$
680 $0.18, p = 0.0036$). Overall accuracy was higher on Ironic than on Literal
681 items ($\beta = 0.6, se = 17, p = 0.0005$). In other terms, in this irony compre-
682 hension task, the rate of false alarms exceeds misses.⁶

683 *Reaction times.* As in Experiment 2, reaction times were recorded from the
684 onset of the word referring to an object in the target until participant's
685 response. Response times, per Group, Type and non-contextual cue are
686 summarised in [Figure 9](#). Linear multilevel regressions, with by-participant
687 random intercepts revealed no effect of Group or of Type ($ps > 0.7$). However,
688 there was an effect of Prosody ($\chi^2(1) = 5.38, p < 0.02$) and Facial expression
689 ($\chi^2(1) = 5.58, p = 0.018$). As predicted, reaction times were shorter in the
690 presence of Prosody ($\beta = -335.1, se = 144, p < 0.02$) and Facial expression
691 ($\beta = -344, se = 145.4, p = 0.018$), across Groups and Type.⁷

⁶The Type X Facial expression interaction it was due to the fact that literal marked
Facial expression improved accuracy on Literal items (viz. reduces the rate of false alarms;
 $\beta = 1.46, se = 0.24, p < 0.0001$).

⁷Using the same method as in Experiment 2 and the same AOIs (see [Figure 7](#)), we
also analysed total fixation durations and fixation counts. Hierarchical multilevel linear

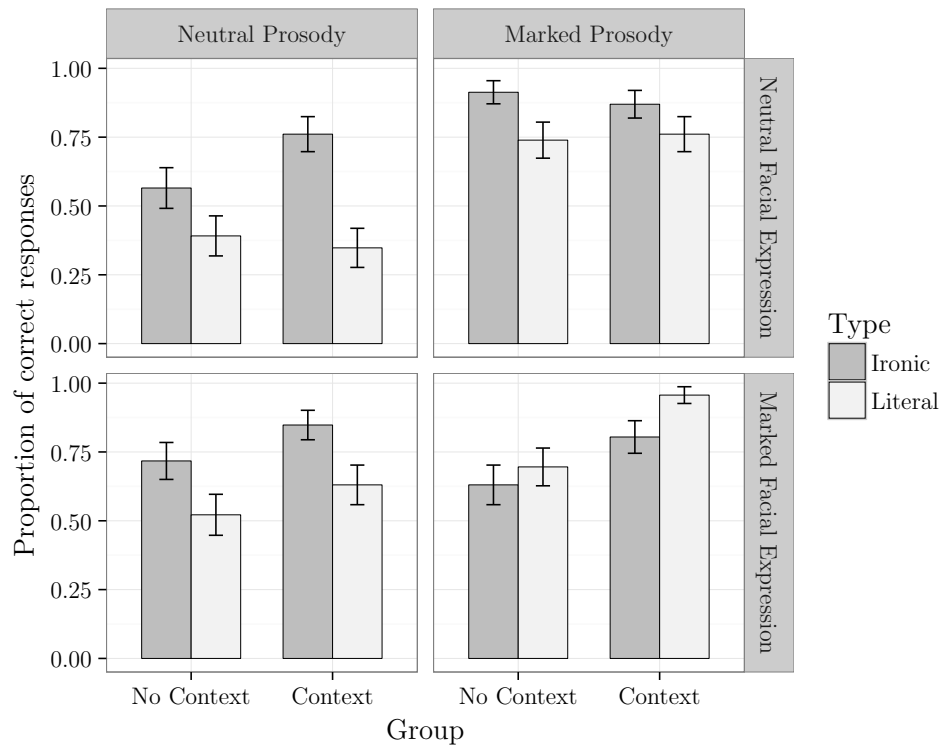


Figure 8: Proportions of correct responses per group, target type and non-contextual cue (vertical bars represent standard errors)

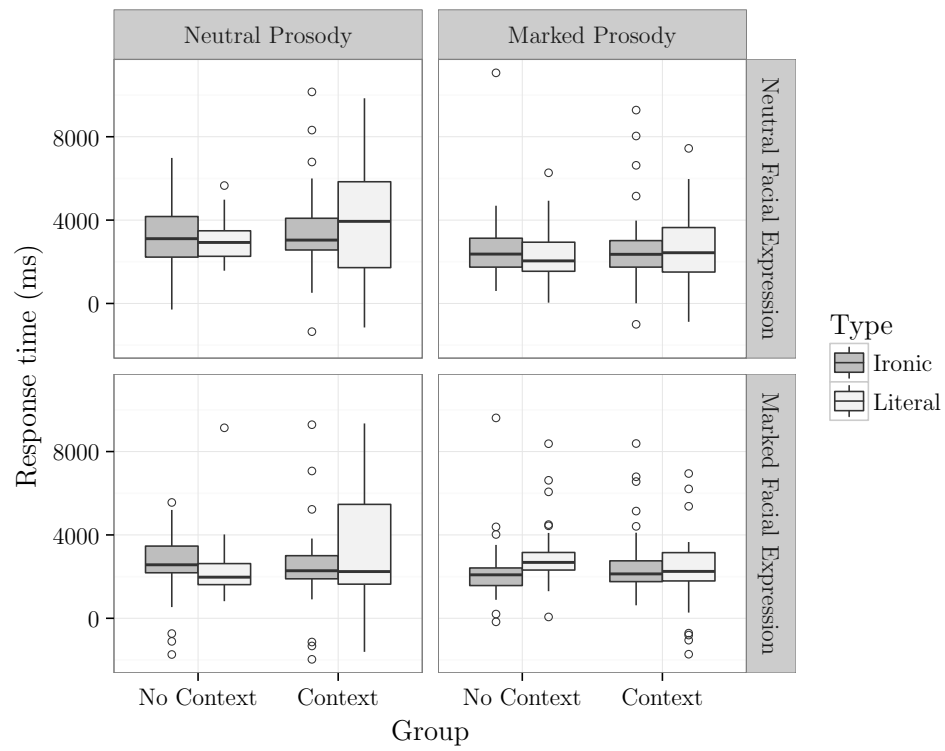


Figure 9: Tukey box-plots for reaction time per group, target type and non-contextual cue

692 *Discussion*

693 Results of our Experiment 3 are entirely consistent with those of [Experiment 2](#), and provide supplementary confirmation for our hypotheses. To
694 begin with, we confirm that contextual incongruence is a much more reliable
695 cue for irony than ironic intonation and facial expression. That is, in line with
696 [Experiment 2](#), high discriminability of these cues, evidenced in [Experiment](#)
697 [1](#), does not translate into comparable reliability in an act-out comprehension
698 task.

700 Furthermore, in Experiment 3 ironic prosody and/or facial expression do
701 not have a cumulative effect with contextual incongruence; if they did, their
702 presence should have entailed higher accuracy in the context group. This
703 result suggests that, as predicted by our [Hypothesis 2](#), intonation and/or fa-
704 cial expression are salient cues that prompt interpreters to terminate costlier
705 context-based processing. This interpretation is reinforced by the fact that,
706 as in [Experiment 2](#), ironic prosody and facial expression are associated with
707 a processing speed-accuracy trade-off. That is, in spite of being less re-
708 liable than contextual incongruence for irony comprehension, the presence
709 of prosody and/or facial expression is associated, in both Context and No
710 Context groups, with shorter response times.

regressions with by-participant random intercept revealed an effect of AOI on total fixation durations ($\chi^2(2) = 113.58, p < 0.001$). However, there is no interaction with Group, Type, Prosody and Facial expression (all $ps > 0.38$). As for fixation counts, there was no effect of AOI ($p = 0.16$). These data are orthogonal to the main point of Experiment 3 and will not be discussed further on.

711 General discussion

712 While ironic prosody and facial cues can be accurately categorized in a
713 discrimination task ([Experiment 1](#)), they do not lead to better grasp of irony
714 in a task where participants must make a decision about the speaker’s com-
715 municative goals ([Experiments 2 and 3](#)). An obvious upshot of our paper,
716 then, is methodological. Researchers should be wary of drawing conclusions
717 about figurative language comprehension based on forced-choice categoriza-
718 tion tasks. The asymmetry between discrimination and use is probably due
719 to the fact that perceptual thresholds between ironic vs. non-ironic prosody
720 and facial expression are not entirely clear-cut. This was made clear by the
721 irony ratings in [Experiment 1](#), which showed that literal positive and neu-
722 tral cues are perceived as more ironic than their negative counterparts. The
723 relative fuzziness of these boundaries has probably less importance in a task
724 where participants have to focus exclusively on locating audio or video stim-
725 uli on an irony scale, but they can lead to more incorrect responses when
726 participants have to make decisions on speaker’s goals. In other terms, cat-
727 egorization of ironic prosody and facial cues can be carried out off-line, but
728 is much more difficult on-line.

729 A potential limitation here could be our use of professional actors in the
730 video stimuli. Although it is a standard practice in the literature on irony
731 (e.g. [Rockwell, 2000](#); [Anolli et al., 2000](#); [Attardo et al., 2003](#); [Rankin et al.,](#)
732 [2009](#)), there is a risk that prosody and facial expression may have been over-
733 played. Recall, however, that results of [Experiment 1](#) did not show any ceiling
734 effect in rating score of prosody and facial expression, and that they led to
735 far from perfect detection of irony in [Experiments 2 and 3](#). To the best of

our knowledge, no study compares prosody and facial expression associated with ironic statements in actors, untrained confederates or in spontaneous speech. One study of acoustic correlates of spontaneous verbal irony reports slower delivery rate as the only robust prosodic characteristic of ironic utterances (Bryant, 2010). A slowdown in speech rate has also been reported in studies using actors (Rockwell, 2000; Anolli et al., 2000), as well as in the current paper. It would be interesting to replicate our findings using recordings of verbal irony in real life situations. However, studying the interplay between ironic cues requires to tightly control the structure of the context segment and the target sentence, which is extremely difficult to achieve in real situations.

The trade-off between accuracy and reaction times, which emerged from Experiments 2 and 3, might look very much like a conceptual conundrum. On the one hand, it seems clear that neither ironic prosody nor ironic facial expression form natural kinds (in line with Bryant & Fox Tree, 2005); on the other hand, participants do seem to privilege such cues, at the expense of accuracy, whenever these are available. On second thought, however, the contradiction is only apparent. Any definition of irony, be it framed in terms of echo or pretence, includes the incompatibility between the context and a literal interpretation of the utterance (Kumon-Nakamura et al., 1995; Wilson, 2006). This is why the capacity to distinguish lies from jokes is operational only if one can make context-based hypotheses about what the speaker wanted the hearer to believe (Wimmer & Leekam, 1991; Martin & McDonald, 2004). It is also for this reason that, as shown by Experiments 2 and 3, assessing the utterance content relative to the background context

761 remains the most reliable route to grasp ironic meanings. In that sense,
762 contextual assessment of the literal meaning is, indeed, an essential part of
763 irony processing, as predicted, for instance, by [Giora \(2003\)](#) and [Sperber &
764 Wilson \(2002\)](#). Yet, even though our results vindicate the central role of
765 context in irony comprehension, they also indicate, consistently with the sec-
766 ond prediction made in the [Introduction](#), that irony processing is not always
767 context-based. One may speculate that along with our communicative expe-
768 rience, grows implicit knowledge that ironic utterances are often accompanied
769 by distinctive prosody or facial expression. Mature communicators may then
770 privilege (what they perceive as) ironic prosody and/or facial expression to
771 speed up the comprehension process. That is, unreliable as they are, these
772 non-contextual cues lead to an activation of ironic meanings without the
773 full-fledged, compositional interpretation being completed.

774 Such a processing route is fully compatible with the Direct Access model
775 ([Gibbs, 2002](#)). It can also be implemented within the parallel-constraint-
776 satisfaction model ([Katz, 2005](#); [Pexman, 2008](#)), provided that this model is
777 amended in way to allow salient non-contextual cues to terminate context-
778 based processing before it is complete. In a way, then, our findings lay ground
779 for reconciling these two models with more context-based theories of irony.
780 It is generally plausible that frugal heuristics are privileged by interpreters
781 whenever possible (in line with, for instance, [Ferreira & Patson 2007](#); [Shintel
782 & Keysar 2009](#); [Epley et al. 2004](#); [Kissine 2016](#)). Assessing the utterance
783 content to the context is a relatively complex, and arguably costly process,
784 so it is not entirely surprising that interpreters forgo it in the presence of
785 more salient cues (as also evidenced by [Deliens et al., 2017](#)).

786 Participants' failure to see that prosody and facial expression are not as
787 reliable as context-based assessment can be profitably conceived of as an in-
788 stance of meta-cognitive error (in the sense of, e.g., [Koriat, 2000](#); [Proust,](#)
789 [2013](#)). According to [Kissine \(2016\)](#), context – understood, this time, as the
790 entire interactional frame of the utterance, including intonation and/or fa-
791 cial expression – plays a two-pronged role in pragmatic processing. On the
792 one hand, it determines the interpretative goal, including, for instance, the
793 level of the specificity of the interpretation output. On the other hand, it
794 is used to monitor and control the interpretation process that has been de-
795 termined by this goal. For instance, the interpretative goal in [Experiment 1](#)
796 consists in mere discrimination of an ironic or not character of a stimulus,
797 which is less complex than genuinely accessing the speaker's intention, as
798 in [Experiments 2](#) and [3](#). Such a superficial ironic interpretation may thus
799 be reached without attempting to assess the speaker's intentions. However,
800 non-contextual processes are less reliable to achieve the more complex in-
801 terpretation goals mandated by the tasks in [Experiments 2](#) and [3](#). Relying
802 on ironic prosody and/or facial expression in these cases, at the expense of
803 context-based processing, thus reflects a meta-cognitive bias, driven by cog-
804 nitive economy principles, which leads participants to select an interpretation
805 process less than optimally suited for the interpretative goal at hand.

806 **Acknowledgements**

807 We are grateful to the journal editor and to the three anonymous reviewers
808 for their detailed and constructive remarks on a previous version of this
809 paper. We also thank the *Jean Raga* theatre company for acting in the

810 video stimuli, and Alba Leszczynski for her help with data collection. At the
 811 time of writing, Gaétane Deliens’s research was supported by a grant from
 812 the Fondation Jean-François Peterbroeck, Kyriakos Antoniou was a Wiener-
 813 Anspach Foundation post-doctoral fellow at the Université libre de Bruxelles.
 814 and Ekaterina Ostaschenko and Fanny Papastamou were supported by the
 815 Research Incentive F.R.S.-FNRS Grant F.4502.15 to Mikhail Kissine. Elise
 816 Clin is a FRESH doctoral fellow at the F.R.S.-FNRS.

817 References

- 818 Akimoto, Y., Miyazawa, S., & Muramoto, T. (2012). Comprehension pro-
 819 cesses of verbal irony: The effects of salience, egocentric context, and
 820 allocentric Theory of Mind. *Metaphor and Symbol*, 27, 217–242.
- 821 Anolli, L., Ciceri, R., & Infantino, M. G. (2000). Irony as a game of implicit-
 822 ness: Acoustic profiles of ironic communication. *Journal of Psycholinguis-
 823 tic Research*, 29, 275–311.
- 824 Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers
 825 of irony and sarcasm. *Humor*, 16, 243–260.
- 826 Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects model-
 827 ing with crossed random effects for subjects and items. *Journal of Memory
 828 and Language*, 59, 390–412.
- 829 Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear
 830 mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.

- 831 Bignell, S., & Cain, K. (2007). Pragmatic aspects of communication and
 832 language comprehension in groups of children differentiated by teacher
 833 ratings of inattention and hyperactivity. *British Journal of Developmental*
 834 *Psychology*, 25, 499–512.
- 835 Boersma, P., & Weenink, D. (2017). *Praat: doing phonetics by computer.*
 836 *Version 6.0.28.* <http://www.praat.org/>.
- 837 Bryant, G. A. (2010). Prosodic contrasts in ironic speech. *Discourse Pro-*
 838 *cesses*, 47, 545–566.
- 839 Bryant, G. A. (2012). Is verbal irony special? *Language and Linguistics*
 840 *Compass*, 6, 673–685.
- 841 Bryant, G. A., & Fox Tree, J. E. (2002). Recognizing verbal irony in spon-
 842 taneous speech. *Metaphor and Symbol*, 17, 99–119.
- 843 Bryant, G. A., & Fox Tree, J. E. (2005). Is there an ironic tone of voice?
 844 *Language and speech*, 48, 257–277.
- 845 Caillies, S., Bertot, V., Motte, J., Raynaud, C., & Abely, M. (2014). Social
 846 cognition in ADHD: irony understanding and recursive theory of mind.
 847 *Research in Developmental Disabilities*, 35, 3191–3198.
- 848 Cheang, H. S., & Pell, M. D. (2011). Recognizing sarcasm without language:
 849 A cross-linguistic study of english and cantonese. *Pragmatics & Cognition*,
 850 19, 203–223.
- 851 Chevallier, C., Noveck, I., Happé, F. G. E., & Wilson, D. (2011). What’s in

- 852 a voice? prosody as a test case for the theory of mind account of autism.
853 *Neuropsychologia*, 49, 507–517.
- 854 Christensen, R. H. B. (2015). ordinal—regression models for or-
855 dinal data. R package version 2015.6-28. [http://www.cran.r-](http://www.cran.r-project.org/package=ordinal/)
856 [project.org/package=ordinal/](http://www.cran.r-project.org/package=ordinal/).
- 857 Climie, E. A., & Pexman, P. M. (2008). Eye gaze provides a window on
858 children’s understanding of verbal irony. *Journal of Cognition and Devel-*
859 *opment*, 9, 257–285.
- 860 Colich, N. L., Wang, A.-T., Rudie, J. D., Hernandez, L. M., Bookheimer,
861 S. Y., & Dapretto, M. (2012). Atypical neural processing of ironic and
862 sincere remarks in children and adolescent with autism spectrum disorders.
863 *Metaphor and Symbol*, 27, 70–92.
- 864 Deliens, G., Antoniu, K., Clin, E., & Kissine, M. (2017). Perspective-taking
865 and frugal strategies: Evidence from sarcasm detection. *Journal of Prag-*
866 *matics*, 119, 33–45.
- 867 Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective
868 taking as egocentric anchoring and adjustment. *Journal of Personality*
869 *and Social Psychology*, 87, 327–39.
- 870 Ferreira, F., & Patson, N. D. (2007). The ‘good enough’ approach to language
871 comprehension. *Language and Linguistics Compass*, 1, 71–83.
- 872 Filippova, E., & Astington, J. W. (2010). Children’s understanding of social-
873 cognitive and social-communicative aspects of discourse irony. *Child De-*
874 *velopment*, 81, 913–928.

- 875 Gerrig, R. J., & Goldvarg, Y. (2000). Additive effects in the perception of
876 sarcasm: Situational disparity and echoic mention. *Metaphor and Symbol*,
877 15, 197–208.
- 878 Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor and Symbol*,
879 15, 5–27.
- 880 Gibbs, R. W. (2002). A new look at literal meaning in understanding what
881 is said and implicated. *Journal of Pragmatics*, 34, 457 – 486.
- 882 Gibbs, R. W., & Colston, H. L. (2012). *Interpreting Figurative Meaning*.
883 Cambridge: Cambridge University Press.
- 884 Giora, R. (2003). *On our Mind. Salience, context, and figurative language*.
885 Oxford: Oxford University Press.
- 886 Giora, R., Givoni, S., & Fein, O. (2015). Defaultness reigns: The case of
887 sarcasm. *Metaphor and Symbol*, 30, 290–313.
- 888 Jacob, H., Kreifelts, B., Nizielski, S., Schutz, A., & Wildgruber, D. (2016).
889 Effects of emotional intelligence on the impression of irony created by the
890 mismatch between verbal and nonverbal cues. *PLoS ONE*, 11, e0163211.
- 891 Katz, A. N. (2005). Discourse and sociocultural factors in understanding
892 nonliteral language. In H. Colston, & A. N. Katz (Eds.) *Figurative language*
893 *comprehension: Social and cultural influences*, (pp. 183–207). Mahwah,
894 NJ: Erlbaum.
- 895 Kissine, M. (2016). Pragmatics as metacognitive control. *Frontiers in Psy-*
896 *chology*, 6.

- 897 Koriat, A. (2000). The Feeling of Knowing: Some Metatheoretical Impli-
898 cations for Consciousness and Control. *Consciousness and Cognition*, *9*,
899 149–171.
- 900 Kowatch, K., Whalen, J. M., & Pexman, P. M. (2013). Irony comprehension
901 in action: A new test of processing for verbal irony. *Discourse Processes*,
902 *50*, 301–315.
- 903 Kreuz, R. J. (2000). The production and processing of verbal irony. *Metaphor*
904 *and Symbol*, *15*, 99–107.
- 905 Kreuz, R. J., & Link, K. E. (2002). Asymmetries in the use of verbal irony.
906 *Journal of Language and Social Psychology*, *21*, 127–143.
- 907 Kreuz, R. J., & Roberts, R. M. (1995). Two cues for verbal irony: Hyperbole
908 and the ironic tone of voice. *Metaphor and Symbol*, *10*, 21–31.
- 909 Kumon-Nakamura, S., Glucksberg, S., & Brown, M. (1995). How about
910 another piece of pie: The allusional pretense theory of discourse irony.
911 *Journal of Experimental Psychology: General*, *124*, 3–21.
- 912 Lenth, R. (2016). Least-squares means: The r package lsmeans. *Journal of*
913 *Statistical Software*, *69*, 1–33.
- 914 Løevenbruck, H., Ben Jannet, M. A., D’Imperio, M., Spini, M., &
915 Champagne-Lavau, M. (2013). Prosodic cues of sarcastic speech in french:
916 slower, higher, wider. In *14th Annual Conference of the International*
917 *Speech Communication Association (Interspeech 2013)*, (pp. 3537–3541).

- 918 Martin, I., & McDonald, S. (2004). An exploration of causes of non-literal lan-
 919 guage problems in individuals with asperger syndrome. *Journal of Autism*
 920 *and Developmental Disorders*, 34, 311–328.
- 921 McDonald, S. (2000). Neuropsychological studies of sarcasm. *Metaphor and*
 922 *Symbol*, 15, 85–98.
- 923 McDonald, S., & Pearce, S. (1996). Clinical insights into pragmatic theory:
 924 Frontal lobe deficits and sarcasm. *Brain Lang*, 53, 81–104.
- 925 Pexman, P. M. (2008). It’s fascinating research the cognition of verbal irony.
 926 *Current Directions in Psychological Science*, 17, 286–290.
- 927 Proust, J. (2013). *The philosophy of metacognition: Mental agency and self-*
 928 *awareness*. Oxford: Oxford University Press.
- 929 R Core Team (2016). *R: A Language and Environment for Statistical Com-*
 930 *puting*. Vienna, Austria: R Foundation for Statistical Computing.
 931 URL <https://www.R-project.org/>
- 932 Rankin, K. P., Salazar, A., Gorno-Tempini, M. L., Sollberger, M., Wilson,
 933 S. M., Pavlic, D., Stanley, C. M., Glenn, S., Weiner, M. W., & Miller,
 934 B. L. (2009). Detecting sarcasm from paralinguistic cues: anatomic and
 935 cognitive correlates in neurodegenerative disease. *Neuroimage*, 47, 2005–
 936 15.
- 937 Rockwell, P. (2000). Lower, slower, louder. vocal cues of sarcasm. *Journal*
 938 *of Psycholinguistic Research*, 29, 483–495.

- 939 Shintel, H., & Keysar, B. (2009). Less is more: A minimalist account of joint
940 action in communication. *Topics in Cognitive Science*, 1, 260–273.
- 941 Sperber, D., & Wilson, D. (1981). Irony and the use-mention distinction.
942 In P. Cole (Ed.) *Radical Pragmatics*, (pp. 295–318). New York: Academic
943 Press.
- 944 Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading.
945 *Mind and Language*, 17, 3–23.
- 946 Spotorno, N., & Noveck, I. A. (2014). When is irony effortful? *Journal of*
947 *Experimental Psychology. General*, 143, 1649–1665.
- 948 Staikova, E., Gomes, H., Tartter, V., McCabe, A., & Halperin, J. M. (2013).
949 Pragmatic deficits and social impairment in children with adhd. *Journal*
950 *of Child Psychology and Psychiatry*, 54, 1275–1283.
- 951 Stalnaker, R. C. (2002). Common ground. *Linguistics and Philosophy*, 25,
952 701–721.
- 953 Voyer, D., Thibodeau, S.-H., & DeLong, B. J. (2016). Context, contrast, and
954 tone of voice in auditory sarcasm perception. *Journal of Psycholinguistic*
955 *Research*, 45, 29–53.
- 956 Wilson, D. (2006). The pragmatics of verbal irony: echo or pretence? *Lingua*,
957 116, 1722–1743.
- 958 Wimmer, E., & Leekam, S. R. (1991). Distinguishing irony from deception:
959 Understanding the speaker’s second-order intention. *British Journal of*
960 *Developmental Psychology*, 9, 257–270.